PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification [4]:
C07H 15/12, C07K 15/00
A61K 39/04, G01N 33/53

A1

(11) International Publication Number: **WO 88/ 06591**

(43) International Publication Date: 7 September 1988 (07.09.88)

(21) International Application Number: PCT/US88/00598

(22) International Filing Date: 25 February 1988 (25.02.88)

(31) Priority Application Numbers: 019,529
Not furnished

(32) Priority Dates: 26 February 1987 (26.02.87)
24 February 1988 (24.02.88)

(33) Priority Country: US

(71) Applicant: SCRIPPS CLINIC AND RESEARCH FOUNDATION [US/US]; 10666 Torrey Pines Road, La Jolla, CA 92037 (US).

(72) Inventors: SHINNICK, Thomas ; 1434 Ranier Falls Drive, Atlanta, GA 30329 (US). HOUGHTEN, Richard ; 358 Ford Avenue, Solana Beach, CA 92075 (US).

(74) Agents: GAMSON, Edward, P.; Dressler, Goldsmith, Shore, Sutker & Milnamow, Ltd., 1800 Prudential Plaza, Chicago, IL 60601 (US) et al.

(81) Designated States: AT (European patent), AU, BE (European patent), CH (European patent), DE (European patent), DK, FI, FR (European patent), GB (European patent), IT (European patent), JP, LU (European patent), NL (European patent), NO, SE (European patent).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: MYCOBACTERIAL RECOMBINANTS AND PEPTIDES

(57) Abstract

Recombinant 540 amino acid residue and 517 amino acid residue proteins encoded by the genome of *Mycobacterium tuberculosis* are disclosed as are vectors for propagating their DNA sequences and expressing the proteins. Also disclosed are methods for using those proteins. Peptides that correspond substantially to the sequences of those proteins and methods of their use are also disclosed, as are polymers containing peptide repeating units corresponding to the 540 residue protein and also polymers containing 517 protein pentapeptides as repeating units.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	ML	Mali
AU	Australia	GA	Gabon	MR	Mauritania
BB	Barbados	GB	United Kingdom	MW	Malawi
BE	Belgium	HU	Hungary	NL	Netherlands
BG	Bulgaria	IT	Italy	NO	Norway
BJ	Benin	JP	Japan	RO	Romania
BR	Brazil	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic			SE	Sweden
CG	Congo	KR	Republic of Korea	SN	Senegal
CH	Switzerland	LI	Liechtenstein	SU	Soviet Union
CM	Cameroon	LK	Sri Lanka	TD	Chad
DE	Germany, Federal Republic of	LU	Luxembourg	TG	Togo
DK	Denmark	MC	Monaco	US	United States of America
FI	Finland	MG	Madagascar		

MYCOBACTERIAL RECOMBINANTS AND PEPTIDES

Description

Cross Reference to a Related Application

This is a continuation-in-part of co-pending application Serial No. 019,529 filed on February 26, 1987, that is incorporated herein by reference.

Technical Field

The present invention relates to recombinant proteins and peptides related to mycobacteria, and particularly to proteins of Mycobacterium tuberculosis that are coded for by adjacent open reading frames on complementary DNA strands of the genome and vectors for propagating and expressing those recombinants, as well as to peptides that correspond substantially in sequence to portions of those proteins.

Background Art

The mycobacteria are a diverse collection of acid-fast, gram-positive bacteria some of which cause important human and animal diseases [reviewed in Bloom et al., (1983), Rev. Infect. Dis., 5:765-780; and Chaparas, (1982), CRC Reviews in Microbiology, 9:139-1971. In man, the two most common mycobacteria-caused diseases are tuberculosis and leprosy, which result from infections with Mycobacterium tuberculosis and Mycobacterium leprae, respectively. These two diseases afflict more than 65 million individuals world-wide and result in over 4 million deaths annually, Bloom et al., (1983), Rev. Infect. Dis., 5:765-780.

The pathogenicity of these mycobacterial infections is closely tied to the host's immune response to the invading mycobacterium [Chaparas,

(1982), CRC Reviews in Microbiology, 9:139-197; Collins, (1982), Am. Rev. Respir. Dis., 125:42-49; Dannenberg, (1982), Am. Rev. Respir. Dis., 125:25-29; and Grange, (1984), Adv. Tuberc. Res., 21:1-78]. Not only does M. tuberculosis infect and grow within cells of the host's immune system, primarily the aveolar macrophage, but also it is the host's cellular immune response that plays the key roles in immunity from infection, containment of the infection at the initial focus of infection, progression or regression of the infection, and tissue damage or destruction at the foci of infection [Chaparas, (1982), CRC Reviews In Microbiology, 9:139-197; Collins, (1982), Am. Rev. Respir. Dis., 125:42-49; Dannenberg, (1982), Am. Rev. Respir. Dis., 125:25-29; and Grange, (1984), Adv. Tuberc. Res., 21:1-78]. In addition, the standard method of detecting an M. tuberculosis infection, the tuberculin skin test, actually measures the host's cellular immune response to the mycobacterium [Snider, (1982), Am. Rev. Respir. Dis., 125:108-118]. The mycobacterial components that are important in eliciting the cellular immune response are not yet well defined.

A number of studies have attempted to define the mycobacterial antigens by standard biochemical and immunological techniques including the analysis of the target antigens of monoclonal hybridoma antibodies directed against mycobacteria [Daniel et al., (1978), Microbiol. Rev., 42:84-113; Engers et al., (1985), Infect. Immun., 48:603-605; Engers et al., (1986), Infect. Immun., 51:718-720; Grange, (1984), Adv. Tuberc. Res., 21:1-78; Ivanyi et al., (1985), Monoclonal Antibodies Against Bacteria (A. J. L. and E. C. Macario, eds.) Academic Press, Inc. New York. pp. 59-90; and Stanford, (1983), The Biology of

the Mycobacteria (Ratledge and Stanford, eds.), Academic Press, London, vol. 2, pp. 85-127].

One particular antigen, a 65 kilodalton (KD) protein, is present in a wide range of mycobacterial species and has been most intensively studied as an antigen of M. leprae [Emmrich et al., (1986), J. Exp. Med., 163:1024-1029; Gillis et al., (1985), Infect. Immun., 49:371-377; Young et al., (1985), Nature, 316:450-452; and Mehra et al., (1986) Proc. Natl. Acad. Sci. USA, 83:7013-7017]. This antigen has been designated the 65KD antigen or the cell wall protein-a (CWP-a) antigen since it appears to a co-purify with cell walls in some isolation procedures [Gillis et al., (1985), Infect. Immun., 49:371-377].

In Western blot assays, monoclonal antibodies directed against this antigen react with two major components in an M. leprae extract that migrate with apparent sizes of 55,000 and 65,000 daltons, and react occasionally with smaller components as well [Engers et al., (1985), Infect. Immun., 48:603-605 and Gillis et al., (1985), Infect. Immun., 37:172-178]. It is not known if these species represent discrete proteins or precursors and products, or result from chemical or enzymatic cleavage during isolation. In other species, such as M. gordonae, only a single species of about 65,000 daltons is detected with the monoclonal antibodies [Gillis et al., (1985), Infect. Immun., 49:371-377].

The 65KD antigen is one of the major immunoreactive proteins of the mycobacteria. This antigen contains epitopes that are unique to a given mycobacterial species as well as epitopes that are shared amongst various species of mycobacteria [Engers et al., (1985), Infect. Immun., 48:603-605

and Gillis et al., (1985), Infect. Immun., 49:371-377]. In addition, some other antigens that appear to be expressed by only one mycobacterial species are also found to contain epitopes expressed in other mycobacterial species. [Kingston et al., (1987) Infect. Immun., 55:3149.]

As discussed hereinafter, it is now found that purified 65KD antigen can elicit a strong delayed-type hypersensitivity reaction in experimental mammals infected with M. tuberculosis. Antibodies directed against this protein can also be detected in the sera of patients with tuberculosis or leprosy, and T-cells reactive with this antigen can be isolated from patients with leprosy or tuberculosis as well as from BCG-vaccinated persons [Emmrich et al., (1986), J. Exp. Med., 163:1024-1029; Engers et al., (1986), Infect. Immun., 51:718-720; Mustafa et al., (1986), Nature, 319:63-66; and Thole et al., (1985), Infect. Immun., 50:800-806]. Overall, the 65KD antigen appears to be a major, medically important B- and T-cell immunogen and antigen in humans.

Brief Summary of the Invention

The present invention relates to DNA sequences, vectors containing the DNA sequences, proteins, recombinant proteins, peptides, their method of manufacture and use that relate to a Mycobacterium tuberculosis. More particularly, those DNA sequences, vectors, proteins, recombinants and peptides relate to two proteins denominated the 540 (65KD) and 517 proteins that are coded for by adjacent open reading frames on complementary DNA strands of the mycobacterial genome. The peptides correspond substantially to portions of those proteins.

One embodiment of the invention contemplates an isolated DNA molecule that consists essentially of a nucleotide sequence, from right to left and in the direction from 5'-end to 3'-end, corresponding to the sequence represented by the formula of Figure 2 from about position 3950 to about position 2390 and in a consistent reading frame coding for a 517 amino acid residue protein of Mycobacterium tuberculosis. More preferably, that sequence extends from position 3948 through position 2398.

A plasmid vector that comprises a replicon operationally linked to a foreign DNA sequence such as that above and that is capable of replicating that foreign DNA sequence in a replication/expression medium is also contemplated herein, particularly where the replication/expression medium is a unicellular organism, such as a bacterium like E. coli. The plasmid vector typically includes sequence-encoded signals for initiation and termination of transcription that are operationally linked to the foreign DNA sequence and are compatible with the replication/expression medium for transcribing a product coded for by the foreign DNA sequence. Further, it can include a translation initiation codon and a translation termination codon, each of which is operationally linked to the 5'-end and the 3'-end, respectively, of the DNA sequence, and are compatible with the replication/expression medium for expressing a protein product coded for by the foreign DNA sequence.

Still further, the 5'-end of the foreign DNA sequence can be operationally linked in translational reading frame to the 3'-end of a second DNA sequence that codes for a second protein or protein fragment or portion, such as the beta-galactosidase molecule.

The protein product expressed by that vector is thus a fusion protein that contains the second protein or protein fragment or portion at the amino-terminus and the first-named protein at the carboxy-terminus of the fusion protein; i.e., the fragment or portion of the second protein is at the amino-terminus of the first-named protein.

A culture comprising bacteria that contain a previously described plasmid vector in an aqueous medium appropriate for the expression of the 517 amino acid residue protein of M. tuberculosis is also contemplated.

The present invention further contemplates a method for producing a 517 amino acid residue protein of M. tuberculosis. That method comprises the steps of culturing a replication/expression medium containing a plasmid vector for replicating and expressing foreign DNA sequence contained therein. That vector contains a foreign DNA sequence that corresponds substantially to the previously mentioned DNA molecule that encodes the sequence of the 517 M. tuberculosis protein. The vector also contains operatively linked nucleotide sequences regulating replication and expression of the foreign DNA sequence. The culturing is carried out under conditions suitable for expression of the protein that is encoded by the foreign DNA. The expressed protein encoded by that foreign DNA sequence is thereafter harvested. Culture is typically carried out using unicellular organisms as the replication/expression medium. Such unicellular organism are typically bacteria as described previously.

A method for determining previous immunological exposure of a mammalian host to

Mycobacterium tuberculosis or Mycobacterium bovis is also contemplated. This method comprises the following steps. An inoculum that consists essentially of the purified 65KD (540) protein or an immunologically active portion thereof coded for by the DNA sequence of Figure 2 is administered intradermally to an assayed mammalian host. That protein is dissolved or dispersed in a physiologically tolerable diluent and is present in that diluent in an amount effective to induce erythema and induration in a mammalian host previously immunized with M. tuberculosis or M. bovis. The mammal is maintained for a time period of about 24 to about 72 hours, and thereafter is assayed for the presence of erythema and induration at the site of the intradermal administration at the end of that time period. In one aspect of this method the purified 65KD protein is obtained from a mycobacterium such as M. tuberculosis. In another aspect of this method, the purified protein is a recombinant 65KD protein, or a recombinant fusion protein that contains a portion of a beta-galactosidase molecule peptide-bonded to the amino-terminus of the 65KD protein, or to the amino-terminus of an immunologically active portion thereof. This type of assay is usually referred to as a delayed cutaneous hypersensitivity (DCH) assay.

Still another aspect of the invention contemplates an inoculum that consists essentially of the purified 65KD (540 amino acid residue) protein antigen or a fusion protein that is coded for by the sequence of Figure 2. That protein antigen is dissolved or dispersed in a physiologically tolerable diluent, and is present in the diluent in an amount that is effective to induce erythema and induration

in a mammalian host previously immunized with M. tuberculosis or M. bovis. The 65KD protein antigen of the inoculum can be one of the proteins useful in the method described immediately above.

Still a further aspect of the invention is a peptide that consists essentially of a 5 to about 40 amino acid residue sequence that corresponds substantially to a sequence of the 540 amino acid residue protein or the 517 amino acid residue protein coded for by the DNA protein sequence of Figure 2. More preferably, the peptide contains about 10 to about 20 amino acid residues.

Preferred peptides include those having a sequence, written from left to right in the direction from amino-terminus to carboxy-terminus using single letter symbols, that corresponds to a formula selected from the group consisting of

A V L E D P Y I L L V S S K V (22; 211-225);
L L V S S K V S T V K D L L P (23; 219-233);
L L P L L E K V I G A G K P L (24; 231-245);
A I L T G G Q V I S E E V G L (30; 291-305);
I A F N S G L E P G V V A E K (46; 451-465);
A R R G L E R G L N A L A D A V K V (58; 11-28);
E K I G A E L V K E V A K K (59; 67-78);
G L K R G I E K A V E K V T E T L (60; 114-130); and
I E D A V R N A K A A V E E G (68; 394-408);

wherein each first parenthesized number refers to the Peptide number of Tables 2 and 4, hereinafter, and the second hyphenated numbers refer to the position in the sequence of the 540 amino acid residue-containing protein whose complete amino acid residue sequence and genomic sequence are illustrated in Figures 2A and 2B.

Further contemplated is a method for ascertaining the presence of mycobacterially-exposed or mycobacterially-immune, i.e., previously immunologically exposed, mononuclear cells such as T cells in a body sample. Here, mononuclear cells from a mammalian host to be assayed are admixed and contacted in an aqueous cell culture medium with a stimulating amount of both antigen presenting cells and a preferred peptide antigen to form a stimulatory cell culture. That stimulatory cell culture is maintained for a time period sufficient for immune mononuclear cells present to be stimulated and to evidence their stimulation. The presence of mononuclear cell stimulation is thereafter determined. This assay can be carried out in vivo as a DCH assay where the antigen presenting cells are endogenous cells such as macrophages and the aqueous medium is supplied by the blood and lymph. The assay can also be carried out in vitro. A polymer having an above peptide as repeating units can also be used as the antigen.

An assay kit containing a preferred peptide in a container in an amount sufficient to carry out at least one assay as described immediately above is also contemplated.

The invention still further contemplates a vaccine against mycobacteria such as M. tuberculosis. The vaccine comprises a physiologically tolerable diluent containing as immunogen an immunizing effective amount of (i) a peptide antigen containing 5 to about 40 residues, and more preferably about 10 to about 20 residues, whose amino acid residue sequence corresponds substantially to a sequence of a mycobacterial 65KD protein and that is capable of stimulating

mycobacterially-immune T cells having a phenotype selected from the group consisting of $T4^+$ and $T8^+$ or (ii) a polymer having said peptide antigen as repeating units. Preferably, the mycobacteria is M. tuberculosis. The mycobacteria to which the T cells are immune is the same mycobacterial species to which the vaccine is directed.

Yet another aspect of the present invention is a polymer that comprises a plurality of pentapeptide repeating units. Each of those pentapeptide repeating units consists essentially of a sequence, written from left to right in the direction of amino-terminus to carboxy-terminus, represented by a formula

N N N I G; or
X G N Z G,

wherein X is an amino acid residue selected from the group consisting of F, S, T, L, D, and I; and Z is an amino acid residue selected from the group consisting of T, I, L, S and V. In a further aspect of this invention, the pentapeptide repeating units are bonded together by peptide bonds, whereas in yet another aspect, the pentapeptide repeating units are bonded together by oxidized cysteine residues at the terminii of those repeating units.

Brief Description of the Drawings

In the drawings forming a portion of this disclosure:

Figure 1 is a schematic restriction map of recombinants expressing the M. tuberculosis 65KD antigen. The portion of the genome containing the 65KD protein is shown as the heavy line at the top of the Figure along with the relative positions (short

perpendicular lines abutting the heavy line) of restriction endonuclease cleavage sites. The single letters adjacent those short lines are indicia of the endonuclease that cleaves the genome at the indicated sites, and are: A = SacI, B = Bgl II, K = KpnI, M = BamHI, P = PstI, R = EcoRI, S = Sal I, V = PvuII, and X = XhoI.

Twenty of the recombinants discussed herein are enumerated along the right-hand margin of the Figure opposite the schematic line representations of the respective genomic portion contained by each recombinant. The lengths and positions of those genomic portions relative to the genome of the 65KD protein are shown by the relative lengths and positions of the lines. Dashes at the termini of the first six shorter lines indicate that those recombinants contained additional base pairs, but the source and sequences of those additional base pairs is presently uncertain.

DNA was isolated from phage stocks of the recombinants expressing the 65KD antigen as described by Helms et al. (1985) DNA 4:39-49, and a restriction enzyme cleavage site map was constructed.

Figure 2 shows the nucleotide sequence of the region containing the M. tuberculosis 65KD antigen and 517 protein genes, and is provided as four sheets labeled 2A, 2B, 2C and 2D. The deduced amino acid residue sequences of the two long open reading frames (ORFs) capable of coding for proteins containing 540 and 517 amino acid residues, respectively, are shown using the one letter code over (540) or under (517) the appropriate triplets. Asterisks above or below the respective sequences indicate the positions of stop codons (TGA, TAG or TAA) in the DNA sequences. Each sequence is shown as

beginning with the first methionine (M) residue in phase with the ORF and downstream of the nearest upstream stop codon.

Figure 3 is a schematic representation of the open reading frames found in the portion of the mycobacterial DNA sequence that codes for the 65KD antigen. The heavy line near the top of the Figure represents a portion of the genome that includes the 540 and 517 proteins. The shorter, arrow-tipped lines beneath the heavy line indicate DNA sequences that exceed 120 amino acid residues in length. Putative initiation triplets are identified on the shorter lines by the letter "M" (AUG) or the letter "V" (GUG) at the 5'-end of each open reading frame in the relatively shorter sequences illustrated beneath the heavy line. Arrows indicate the coding direction.

Figure 4 is a photograph of a Western blot analysis of products of the 540 amino acid residue open reading frame, and contains two panels, A and B. Cells were grown and induced (except for lane 2, Panel A) and crude extracts were prepared as described in the Materials and Methods section, hereinafter. For each lane, except lane 5, 200 micrograms (ug) of protein were electrophoresed on a 10% Laemmli gel, and transferred to nitrocellulose. For lane 5, 500 ug of protein were loaded. The immobilized proteins were reacted with the IT-13 antibodies and visualized, as discussed hereinafter.

For Panel A, the proteins in the lanes were: lane 1, JM83; lane 2, JM83 (pTB22) uninduced; and lane 3, JM83 (pTB22) induced with IPTG. For Panel B, the proteins in the lanes were: lane 1, JM83 (pTB12); lane 2, Y1089 (ΔSK116); lane 3, Y1089

(λ RY3146); lane 4, BNN97 (*E. coli* C600 containing λ gt11); and lane 5, JM83 (pTB12).

DEFINITIONS

The following abbreviations and symbols are used herein.

bp	-	base pair(s)
kbp	-	1000 bp
KD	-	kilodalton(s)
M_r	-	apparent relative molecular mass
DNA	-	deoxyribonucleic acid
replicon	-	the unit that controls individual acts of replication; it has an origin at which replication is initiated and it can have a terminus at which replication stops.

When used in a context describing or depicting nucleotide sequences, the purine or pyrimidine bases forming the nucleotide sequence are depicted as follows:

A	-	deoxyadenyl
G	-	deoxyguanyl
C	-	deoxycytosyl
T	-	deoxythymidyl

In describing a nucleotide sequence each three-letter triplet constituted by the bases identified above represents a trinucleotide of DNA (a codon) having a 5'-end on the left and a 3'-end on the right of the upper sequence of Figure 2, and a 5'-end on the right and a 3'-end on the left of the lower, complementary sequence.

The word "antigen" is often used in the art for an entity that is bound by an antibody. The word "immunogen" is often used in the same context for the entity that induces the production of antibodies. Where the antigen and immunogen are the same entity, both are often referred to an antigen.

The present invention deals with antigens and immunogens in the above context, which context typically relates to B cells and antibodies. Notwithstanding the B cell/antibody context, the present invention also contemplates T cells.

A more general definition of immunogen and antigen apply in the context of T cells and T cell stimulation. In that more general definition, an "antigen" is an entity acted upon by a component of the immune system, and an "immunogen" is an entity that initiates an immune system response. Where antigen and immunogen are the same, both are referred to as an antigen. An "immunologically active" entity interacts with antibodies or T cells, or can initiate a cellular or humoral immune response.

Detailed Description of the Invention

I. OVERVIEW

In studies discussed hereinafter, the isolation of the gene encoding the M. tuberculosis 65KD antigen and the determination of its nucleotide sequence are reported. The sequence contains an open reading frame encoding 540 amino acid residues or about 60,000 daltons, which corresponds to the 65KD antigen. A second long open reading frame capable of encoding a protein of 517 amino acids was also found on the mycobacterial DNA fragment containing the 65KD antigen gene, adjacent to that gene. Interestingly, the central region of the deduced amino acid residue sequence of the 517 amino acid protein contains

several tandemly arranged, perfect and imperfect repeats of a five amino acid residue sequence. This feature is reminiscient of the features of the sequence of the major T-cell antigen of the sporozoite stage of the human malarial parasite [Nussenzweig et al., (1985), Cell, 42:401-403].

II. RESULTS

A. Isolation and Analysis of Recombinants Expressing the 65KD Antigen

To isolate the gene that encodes the 65KD antigen, monoclonal hybridoma antibodies directed against this antigen were used to screen a protein expression library constructed with mycobacterial DNA. An expression library was chosen since it was not known *a priori* if the *M. tuberculosis* genes would be expressed in *E. coli*. Such a recombinant DNA library has been constructed by Young et al., (1985), Proc. Natl. Acad. Sci. USA, 82:2583-2587, and contains genomic DNA fragments of *M. tuberculosis* inserted into the expression site of the lambda-gt11 (λgt11) vector. In this system, the inserted coding sequences can be expressed as a fusion protein with beta-galactosidase. The 65KD antigen-specific monoclonal hybridoma antibodies used in these studies were generated in the laboratories of Dr. T. M. Buchanon (Pacific Medical Center, University of Washinton, Seattle WA) and Dr. J. Ivanyi (MRC Tuberculosis Unit, Hammersmith Hospital, London) and were obtained from the Steering Committee on the Immunology of Tuberculosis of the World Health Organization.

As the initial antibody probe, a pool containing three monoclonal antibodies directed against the 65KD antigen was used (IT-13, IT-31, and IT-33). Thirty-eight positive signals were detected in a screen of about 8x105 recombinant phage.

The phage corresponding to the positive signals were twice plaque purified and then assayed for reactivity with the individual antibodies. The results of that purification and assay are shown in Table 1, below.

TABLE 1

Patterns of Antibody Reactivities[1]

	Reactivity With Antibodies		
Number of Clones	IT-13	IT-31	IT-33
28	+	+	+
3	+	+	-
3	-	+	+
2	-	+	-
2	-	-	+

[1]Recombinant clones expressing antigens reactive with the 65KD antigen specific monoclonal antibodies IT-13, IT-31, and IT-33 were isolated as described in the text. For the initial screen, a pool of the three antibodies that contained a 1:1000 dilution of each antibody was used to screen a total of about 8×10^5 recombinant phage from the lambda gt11-*M. tuberculosis* library. To determine which monoclonal antibody reacted with which of the 38 plaque-purified recombinants, about 100 plaque-forming units (pfu) of each recombinant phage were inoculated in small spots on a lawn of *E. coli* Y1090. The phage were allowed to grow, and were induced to synthesize the foreign proteins as described herein. The filters were then reacted with a 1:1000 dilution of one of the monoclonal hybridoma antibodies as described in Materials and Methods.

Twenty-eight of the recombinants produced antigens that reacted with all three antibodies, whereas ten recombinants produced antigens that reacted with one or two of the antibodies. Overall, the patterns of reactivity indicate that although the three antibodies react with the same mycobacterial antigen, each recognizes a different epitope on that antigen. Richard A. Young (Whitehead Institute, M.I.T.) has also screened this λgt11-*M. tuberculosis* library with one of these antibodies (IT-13) and detected 10 additional recombinants [Young et al., (1985), Proc. Natl. Acad. Sci. USA, 82:2583-258]. These recombinants were not assayed for reactivity with the other antibodies.

DNA was isolated from twenty of the recombinants expressing the 65KD antigen and a restriction enzyme cleavage site map was deduced for this region of the mycobacterial genome (Figure 1). In most of the recombinants, the mycobacterial DNA insert was flanked by EcoRI sites as expected from the way in which the library was constructed.

However, in 6 of the 20 recombinants studied, only one of the expected EcoRI sites was present. This observation raises the possibility that a significant fraction of the recombinant phage in this library might have arisen from the insertion of a fragment containing only one functional EcoRI site into the λgt11 EcoRI site or that some clones might have undergone some sort of recombination, rearrangement or deletion event during propagation that removed one of the EcoRI sites.

The deduced restriction map is in good agreement with the published map of the gene for the *M. bovis* 65KD antigen [Thole et al., (1985), Infect. Immun., 50:800-806] except for the presence of two

additional SmaI sites in the M. tuberculosis gene. The map does not match well with that of the M. leprae 65KD antigen gene [Young et al., (1985), Nature, 316:450-452]. This is not unexpected given that based on DNA homology studies, M. tuberculosis is at least 90% homologous with M. bovis and only about 30% homologous with M. leprae, Athway et al., (1984), Int. J. Syst. Bacteriol., 34:371-375; Imaeda, (1985) Int. J. Syst. Bacteriol., 35:147-150.

To determine the nucleotide sequence of this region of the mycobacterial genome, several fragments from the λgt11 recombinants were subcloned into the plasmid vector pUC19. The majority of the sequence of this region was determined from a subclone (pTB7) of the 1.4 kilobase pair (kbp) EcoRI fragment of λSK7 and a subclone (pTB9) of the 2.6 kbp EcoRI fragment of λRY3143. The sequence across the EcoRI site at the junction of these two fragments was determined from a fragment isolated from a subclone (pTB11) of the 2.8 kbp KpnI fragment of λSK119. The sequence of the region 5' to the 2.6 kbp EcoRI fragment was determined from a subclone (pTB12) of the 2.4 kbp KpnI fragment of λSK119.

In all, the nucleotide sequence of 4380 base pairs of the mycobacterial DNA was determined by a combination of the Sanger dideoxy chain termination [Sanger et al., (1980), J. Mol. Biol., 143:161-178] and Maxam-Gilbert chemical degradation [Maxam et al., (1976), Proc. Natl. Acad. Sci. USA, 74:560-564] sequencing techniques. The sequence is shown in Figure 2.

As expected for M. tuberculosis genomic DNA [Wayne et al., (1968), J. Bacteriol., 96:1916-1919], the base composition of this fragment was about 66% G+C. The high G+C content increased the chances of

sequencing artifacts due to compressions, and made it imperative that the sequences were determined for both strands in all regions.

B. Open Reading Frames

The sequence contains five open reading frames (ORFs) that begin with an ATG triplet and contain greater than 120 amino acids. Two of these exceed 200 amino acids in length. One can encode 517 amino acids and the other 340 amino acids.

There are an additional three open reading frames of 140-190 amino acid residues in length that do not contain an initiation ATG triplet but do contain a GTG triplet. It is not known if a GTG triplet can function as a translation initiation triplet in mycobacteria. The locations of these eight open reading frames are shown schematically in Figure 3. No portions of the deduced amino acid sequences of any of these open reading frames displayed any significant homologies with sequences in the Protein Sequence Database of the Protein Identification Resource.

It should be noted that although an open reading frame exceeding 100 amino acids would be considered to have a high probability of being expressed into protein in most bacteria, this may not be true for the mycobacteria. That is, given that the G+C content of the insert is about 66%, a translation termination triplet (TAA, TAG or TGA) would be expected to occur on average about once every 41 amino acids as compared to about once every 21 amino acids in a genome with a G+C content of 50%. Perhaps then, an open reading frame of as many as 150-200 amino acids might be due to the random distribution of termination triplets rather than signifying possible biologic importance. As such,

only the two very long open reading frames that could encode proteins of 517 and 540 amino acid residues, respectively, are described herein.

C. The 540 Amino Acid Residue ORF Corresponds to the 65KD Antigen

One of the long open reading frames begins with an ATG triplet at positions 252-254 of the DNA sequence and extends to a TGA triplet at positions 1872-1874. This ORF encodes 540 amino acids. To determine if this open reading frame corresponded to the gene for the 65KD antigen, the 1511 bp BamHI-KpnI fragment from pTB12 (residues 438-1948 of the sequence represented in Figure 2), which contains the majority of this open reading frame, was inserted into BamHI-KpnI-cleaved pUC19. In this construct, denominated pTB22, the open reading frame is expressed using the lacZ transcription and translation initiation signals present in the pUC19 vector, and results in the production of a fusion protein containing 15 amino acid residues at the amino-terminus encoded by the lacZ gene of pUC19 followed by 478 amino acids of the mycobacterial open reading frame.

Crude extracts were prepared from cells containing this plasmid, and were tested for reactivity with 65KD antigen-specific antibodies in Western blot analyses. The reactivity with monoclonal antibody IT-13 is shown in panel A of Figure 4. In all, five different monoclonal antibodies specific for the 65KD antigen reacted with a species in the crude extract that migrated with an apparent relative molecular mass (M_r) of about 55,000 daltons (lane 3).

No reactivity was seen in extracts of E. coli lacking the plasmid (lane 1). Furthermore,

the expression of this fusion protein is inducible with isopropyl-beta-D-thiogalactopyranoside (IPTG) (compare lanes 2 and 3). Therefore, it is concluded that this long open reading frame encompassing residues 252-1871 encodes the M. tuberculosis 65KD antigen. The phrases "540 amino acid residue protein", "540 protein", "65KD protein" and "65KD protein antigen" are used interchangeably herein for the 65KD protein of M. tuberculosis.

In addition, the purified recombinant 65KD protein was used in Western blot analyses using serum from human patients known to be infected with M. tuberculosis. In preliminary studies, antisera from those patients immunoreacted with the purified recombinant protein.

Those studies illustrate the use of that natural or recombinant protein as an antigen in a diagnostic assay method for the presence of naturally occurring antibodies to the 65KD protein in the infected patients, and thus for the detection of a Mycobacterium tuberculosis infection in those patients. Similar results are obtained in a more usual solid phase assay such as are carried out in a microtiter plate where the recombinant 65KD protein is affixed to a solid phase matrix to form a solid phase support and patient serum is the source of antibodies to be assayed.

Solid phase assays whether carried out in a microtiter plate, a dipstick or as a Western blot all require the similar steps and constitute variants of each other. Each has a solid phase matrix (mirotiter plate well, stick surface or nitrocellulose) to which the purified natural or a recombinant 540 amino acid protein coded for by the genome of M. tuberculosis as antigen is affixed, usually by adsorption, to form

the solid phase support. The assayed sample such as patient serum or cerebrospinal fluid (where evidence of tubucular meningitis is sought to be assayed) in liquid form is admixed with the solid phase support to form a solid-liquid phase admixture. That admixture is maintained under usual biological assay conditions (e.g. zero degrees C to about 40 degrees C) for a time period sufficient for any antibodies present in the assayed sample to immunoreact with and bind to the antigen of the solid phase support. The solid and liquid phases are separated as by rinsing. The presence of antibodies bound to the solid support is thereafter determined as with a labeled reagent that reacts with the bound human antibodies.

A labeled reagent that reacts with bound human antibodies present is admixed with the solid phase to form a second solid-liquid phase admixture. That second solid-liquid phase admixture is maintained for a time period sufficient for the labeled reagent to react with the bound human antibodies. The second solid-liquid phase admixture is separated as by rinsing, and the amount of label present is determined. An amount of label present above a background, control value indicates the presence of anti-65KD protein antibodies and thus an infection by M. tuberculosis.

The labeled reagent that reacts with the bound human antibodies is preferably a labeled preparation of xenogenic anti-human antibodies such as alkaline phosphatase-conjugated goat anti-human Ig antibodies that are available from Tago, Burlingame, CA. The presence of the bound alkaline phosphatase is typically determined spectrophotometrically by measurement of the enzymatic hydrolysis of a substrate molecule such as p-nitrophenyl phosphate to

p-nitrophenol. Other enzymes such as horseradish peroxidase and other label types such as radioactive elements like iodine 125 are also useful. S. aureus protein A linked to a label such as ^{125}I can also react with the bound human antibodies of the separated solid phases to detect their presence.

The above diagnostic assay method is typically carried out in a clinical setting using a kit. The kit comprises at least one package that contains a solid phase support having a purified 540 protein encoded by the M. tuberculosis genome that is from the mycobacterium or is a recombinant protein as discussed herein affixed as an antigen to a solid matrix such as a plastic microtiter plate or dipstick. One or more additional reagents such as the labeled reagent that reacts with solid phase-bound human antibodies, a substrate for the labeled reagent (where needed for the label), buffer salts in solution or dry form, and the like can also be present in separate packages in the kit.

D. The 65KD Antigen Gene is Expressed in E. coli

Because previous studies had shown that most mycobacterial genes were not expressed in E. coli using the mycobacterial transcription and translation signal sequences [Clark-Curtis et al., (1985), J. Bacteriol., 161:1093-1102; and Thole et al., (1985), Infect. Immun., 50:800-806] a protein expression library was used in the cloning studies. In the λ gtll-M. tuberculosis library, the inserted mycobacterial coding sequences should be expressed as fusion proteins with beta-galactosidase [Young et al., (1983) Proc. Natl. Acad. Sci. USA, 82:2583-2587]. It was somewhat surprising to find that the open reading frame encoding the 65KD antigen

did not extend to the 5'-end of the mycobacterial DNA insert in λSK119. This suggested that the 65KD antigen was being expressed using the mycobacterial transcription and translation signal sequences.

With respect to the previously described E. coli consensus signal sequences, the mycobacterial sequences 180-230 base pairs upstream of the presumed initiator ATG codon do display reasonable matches with the consensus sequences for the -35 (3/3 match with the highly conserved TTG) and -10 (4/6 match with TATAAT) regions of E. coli promoters [Rosenberg et al., (1979), Ann. Rev. Genet., 13:319-353]. There is also a 5/5 match with the Shine-Dalgarno sequence [Shine et al., (1974), Proc. Natl. Acad. Sci. USA, 71:1342-1346] for a prokaryotic ribosome binding site (GGAGG) 13 base pairs upstream of the presumed initiator triplet for the 65KD antigen open reading frame. Although the precise locations of the mycobacterial regulatory sequences have not been determined experimentally, the results of the two studies described below suggest that the mycobacterial sequences are indeed functional in E. coli.

The size of the anti-65KD reactive material produced by the recombinants was determined in a Western blot assay. To do this, crude lysates of cells expressing recombinant plasmids or phage that had been shown to contain the entire 65KD antigen gene (λSK116, pTB12) as well as those that had been shown to contain a large portion of the 65KD antigen open reading frame fused to B-galactosidase (λRY3146; pTB22 that contains the 540 protein DNA from position 438 through position 1948 of Figure 2) were prepared as described in the Materials and Methods section.

The lysates were electrophoresed on 10% Laemmli SDS-polyacrylamide gels, and the separated proteins were electrophoretically transferred to nitrocellulose. The SDS-denatured, immobilized proteins were then reacted with monoclonal antibodies specific for the 65KD antigen.

The results using antibody IT-13 are shown in Figure 4. In cells expressing recombinants carrying the fused open reading frame, the monoclonal antibodies detected a single strongly reactive species migrating with an M_r of about 160,000 daltons as well as occasionally detecting smaller species (Figure 4, Panel B, lane 3). In another fused open reading frame recombinant, the monoclonal antibodies detected a single reactive species migrating with an M_r of about 55,000 daltons (Figure 4, Panel A, lane 3). In the extracts of the cells expressing recombinants that contained the entire 65KD gene, the monoclonal antibodies detected a single strongly reactive species that migrated with an M_r of about 64,000 daltons (Figure 4, Panel B, lanes 1 and 2).

Smaller reacting species (about 40,000-55,000 daltons) were observed when large amounts of the extracts were loaded (lane 5) or when the protease inhibitor was omitted from the lysis buffer. Occasionally, a minor reacting species was also observed migrating with an M_r of about 67,000 daltons.

Given the sizes of the anti-65KD-reactive materials, these data indicate that the 65KD antigen can be expressed using the mycobacterial translation initiation signals present in the 65KD gene. Also, since the vector contribution to the recombinant plasmids does not contain any known sequences that

are properly located and oriented to promote the transcription of the inserted DNA, these data suggest that the mycobacterial transcription initiation signals function in E. coli to allow the expression of the 65KD antigen.

In order to obtain an approximate measure of the efficiency of utilization of the mycobacterial transcription and translation initiation signals in E. coli, two plasmids were constructed that placed the expression of enzymatically active beta-galactosidase under the control of either the mycobacterial signal sequences or the lac gene signal sequences present in the plasmid pUC19.

First, the 3000 bp BamHI fragment from pMC1871 that contains the coding sequences for amino acid residues 8-1021 of beta-galactosidase [Shapira et al., (1983), Gene, 25:71-82] was inserted into the BamHI site of pTB12 (residues 437-442 of the sequence presented in Figure 2). The resulting 8.1 kbp plasmid (pTB27) contains an open reading frame that encodes a fusion protein with 63 amino acid residues derived from the 65KD antigen gene followed by 1014 amino acids of beta-galactosidase, and whose expression is under the control of the transcription and translation signal sequences present in the mycobacterial DNA. As expected, this construct expresses a protein of about 120,000 daltons that reacted with anti-beta-galactosidase antibodies in a Western blot assay.

Second, the 3000 bp BamHI fragment from pMC1871 was inserted into the BamHI site in the polylinker of pTB9 that contains a 2.4 kbp fragment of the 65KD antigen gene inserted in the EcoRI site of pUC19. The resulting 8.1 kbp plasmid (pTB28) contains an open reading frame that encodes a fusion

protein with 15 amino acid residues derived from the pUC19 lacZ gene and polylinker sequences followed by the 1014 amino acid residues of beta-galactosidase and whose expression is under the control of the lac gene signal sequences present in pUC19.

Crude extracts of cells containing these plasmids were assayed for beta-galactosidase activity as previously described. In cells containing pTB27, beta-galactosidase activity [about 2800 units/microgram (ug) protein] was about one-fourth that (11,000 units/ug protein) found in IPTG-induced cells containing pTB28. Given the unknowns inherent in this study (e.g., the specific activities and relative stabilities of the two fusion proteins), one cannot make a precise quantitative statement about the relative strengths of the mycobacterial signal sequences and the E. coli lac gene signal sequences based on the relative enzymatic activities found in the two cell extracts. However, the data do indicate that these mycobacterial transcription and translation signal sequences are efficiently recognized in E. coli.

E. The 65KD Antigen Sequence

Several interesting features of this long open reading frame have been revealed by a computer-aided analysis of the sequence. The overall base composition of this open reading frame is 65.5% G+C. However, the G+C content varies considerably within the codons such that the G+C content of the bases occupying the first two residues of the codons is 55% while it is 87% for the bases found in the third position of the codons; thereby producing a bias towards using codons that have a G or C in the third position.

For example, 50 of the 51 leucine codons (CTX) have a G or C in the third position. Interestingly, the essentially random occurence of any of the four bases in the first two positions of a codon plus the preference for G or C in the third position of a codon is one strategy that allows an organism to have a high G+C content without limiting access to the amino acids whose codons contain A or T residues in the first two positions.

Although the deduced amino acid residue sequence of the 65KD antigen is particularly rich in alanine, glycine, leucine, and valine residues, the overall amino acid residue composition contains 52% hydrophobic and 48% hydrophilic residues. Computer-aided analysis of the alpha helical content Chou et al., (1978), Adv. Enzym., 47:45-148 and hydrophobicity (Hopp et al., (1981), Proc. Natl. Acad. Sci. USA, 78:3824-3828) of the amino acid residue sequence revealed numerous regions that could participate in alpha helical structures and no extended regions of high hydrophobicity. These data suggest that the 65KD antigen is not an integral membrane protein but rather its sequence resembles that of a soluble protein.

As discussed before, the 65KD antigen appears to be a major T cell immunogen and antigen in man. It has been suggested that immunodominant T cell epitopes are short stretches of amino acids that can form amphiphilic helices where one side of the helix is hydrophobic and the other side hydrophilic, Berzofsky, (1985), Science, 229:932-940. Based on computer modeling, seven stretches of amino acids within the sequence of the 65KD antigen have been identified that could form such amphiphilic helices. A list of those peptides is shown in Table 2, below.

TABLE 2

Residue Positions[1]	Sequence[2]
11-28 (58)	A R R G L E R G L N A L A D A V K V
66-79 (59)	E K I G A E L V K E V A K K
114-130 (60)	G L K R G I E K A V E K V T E T L
154-172 (61)	Q S I G D L I A E A M D K V G N E G V
219-233 (23)	L L V S S K V S T V K D L L P
394-408 (62)	I E D A V R N A K A A V E E G
494-508 (63)	V K V T R S A L Q N A A S I A

[1]Residue positions are denominated using the one letter amino residue sequence of the 65KD protein shown in Figure 2 that depicts the methionine residue coded for by the triplet beginning at base pair position 252 as the first residue of the protein. Parenthesized numbers refer to peptide numbers that begin with petide number 1 shown in Table 4.

[2]These amino acid sequences are shown from left to right and in the direction from amino-terminus to carboxy-terminus, as is customary in the art.

F. DCH Assay With A Recombinant 65KD Protein

Exemplary delayed cutaneous hypersensitivity (DCH) assays were carried out using illustrative

recombinant proteins described herein as test antigens after immunization with M. tuberculosis, M. bovis or saline. These assays were carried out following the procedure described in Minden et al. (1986) Infec. Immun. 53:560-564.

Briefly, the mammalian hosts were immunized with a sufficient amount of M. tuberculosis or M. bovis to induce an immunological response, or with a control (saline). After maintaining the animals for a time period sufficient for the initial immunological response to the immunogen to subside, the animals were challenged by intradermal injection with inocula containing the 65KD protein, a recombinant 65KD protein, or a recombinant fusion protein that contained the 65KD protein as the test antigen dissolved or dispersed in a physiologically tolerable diluent, or with a control. The test antigens were present in an amount sufficient to induce erythema and induration at the site of administration in a mammal previously immunized with M. tuberculosis or M. bovis.

The results of this study are shown in Table 3, below.

Table 3

DCH Assays With Recombinant Antigens

Challenge Antigen[1]		No. Positive/No. Assayed Of Guinea Pigs Immunized With[2]: M. tuberculosis	M. bovis	Saline
Saline	(0)	0/5	0/5	0/5
BNN97[3]	(10)	0/5	0/5	0/5

λ 1089[4]	(10)	5/5	5/5	0/5
λ1089[4]	(1)	5/5	5/5	0/5
pTB22[5]	(10)	5/5	5/5	0/5
pTB22[5]	(1)	5/5	5/5	0/5
BCG-S[6]	(1)	5/5	5/5	0/5
PPd[7]	(5 T.U.)	5/5	5/5	0/5

[1]Challenge antigen compositions were injected intradermally as discussed in Materials and Methods using amounts of 1 or 10 ug/100 ul per injection as indicated by the parenthesized numeral after each antigen, except for purified protein derivative (PPd) that was used in an amount of 5 tuberculin units (T.U.).

[2]The number of guinea pigs exhibiting positive DCH responses is in the numerator, whereas the number of guinea pigs assayed is in the denominator. The immunization protocol is described in Materials and Methods.

[3]BNN97 was a crude lysate prepared from λgt11-infected E. coli. The crude lysate was partially purified by ammonium sulfate precipitation as described in the Materials and Methods section.

[4]λ1089 was a crude lysate prepared from λSK119-infected E. coli that expressed the 65KD antigen. The crude lysate was partially purified by ammonium sulfate precipitation as described in the Materials and Methods section.

[5]pTB22 was a crude lysate prepared from E. coli containing pTB22 that expressed the 65KD antigen as a fusion protein that contained a portion of the beta-galactosidase molecule and about the carboxy-terminal 88 percent of the 65KD protein. The crude lysate was partially purified by ammonium sulfate precipitation as described in the Materials and Methods section.

[6]BCG-S was an extract of M. tuberculosis prepared as described in the Materials and Methods section.

[7]ppd was obtained from Connaught Laboratories, Ltd., Willowdale, Ontario, Canada.

As can be seen from the above results, the 65KD protein coded for by the DNA sequence of Figure 2 can be utilized in DCH as part of a method to determine whether a mammalian host such as guinea pig had previous immunolgical exposure to M. tuberculosis since the T leucocytes of the host animals produced erythema and induration at the sites of administration in the animals previously immunized with M. tuberculosis and M. bovis, and produced no reactions in the saline-immunized animals. Those results also show that recombinant 65KD protein molecules are similarly useful. Recombinant fusion proteins that contain a portion of the beta-galactosidase molecule peptide-bonded to the amino-terminus of the 65KD protein are also useful, as are fusion proteins that contain a portion of the beta-galactosidase molecule and an immunologically active portion, about the carboxy-terminal 85% of the 65KD protein, e.g., the protein expressed by pTB22.

Fusion proteins that contain one or more peptide sequences as are described in Tables 2 and 4 hereinafter are also useful. The phrase "previous immunological exposure" and its grammatical variants is used herein to mean that the mammalian host had been immunized or infected by one of the mycobacteria and the host mammal mounted an immune response (primary response) to the immunogens provided by the mycobacteria, and that that immune response had subsided.

G. The 517 Amino Acid Protein

1. The Open Reading Frame

A second long open reading frame begins with an ATG codon at positions 3948-3946 of Figure 2 and extends to a TAA triplet at positions 2397-2395 on the DNA strand complementary to the DNA strand encoding the 65KD antigen, thereby making those open reading frames adjacent in the genome. This open reading frame can encode a protein that contains a sequence of 517 amino acid residues, and that protein is referred to herein as the "517 amino acid protein" or the "517 protein". The 517 protein coding region thus extends from position 3948 through position 2398 of Figure 2.

Given that the two long open reading frames are located adjacent and downstream from each other on the complementary strands, one might expect that the transcription of one gene might interfere with the transcription of the other unless there were transcription termination signals within the intergenic region. Indeed, there are several short sequences (e.g., 2134-2160) within the 520 base pair intergenic region that have features reminiscient of the transcription termination signals of gram-negative bacteria [Rosenberg et al., (1979),

Ann. Rev. Genet., 13:319-353]. That is, regions containing short, G+C-rich, inverted repeats capable of forming stem and loop structures followed by a stretch of three or more T residues about 20 bases from the center of dyad symmetry. Perhaps these inverted repeats might function as transcription termination signals to allow the independent expression of each of these mycobacterial genes.

To determine if the 517 amino acid open reading frame was expressed into protein in E. coli, extracts of cells containing a plasmid (pTB11) carrying the complete open reading frame were probed with a polyclonal rabbit antiserum elicited with a sonicated extract of M. tuberculosis bacteria in a Western blot assay. In these recombinants, the putative protein product of the 517 amino acid open reading frame would have to be expressed using the mycobacterial regulatory sequences. The polyclonal antiserum detected more than 100 species in an extract of M. tuberculosis cells as well as the 65KD antigen in extracts of E. coli cells carrying the appropriate plasmid (pTB12), but did not detect any novel proteins in extracts of E. coli cells containing plasmids carrying the 517 amino acid residue protein open reading frame. Hence, either this open reading frame is not expressed in E. coli using the mycobacterial regulatory sequences or the particular antiserum used in the immunoblots did not contain antibodies directed against this protein.

It is not surprising that this open reading frame is not expressed in E. coli using the before-discussed recombinant since previous studies suggest that most mycobacterial genes are not expressed in E. coli [Clark-Curtiss et al., (1985), J. Bacteriol., 161:1093-1102; and Thole et al.,

(1985), Infect. Immun., 50-800-806]. Also, this open reading frame does not contain any impressive matches to the E. coli consensus promoter sequences within the 400 bases upstream of the ATG triplet although it does contain a 3/5 match with the Shine-Dalgarno consensus sequence for ribosome binding sites 12 bases upstream of the initiator ATG triplet. Nonetheless, given the size of this open reading frame and its unique structural features (discussed below), it most likely is expressed into protein in M. tuberculosis and can be expressed in E. coli using a recombinant vector designed for that expression, as is discussed hereinafter.

2. Structural Features of the 517 Protein

The second long open reading frame could encode 517 amino acids or a protein of about 51,000 daltons (calculated M.W.=50,561). The deduced amino acid residue sequence is rich in alanine, asparagine, glycine, and serine and overall is composed of 54% hydrophobic residues and 46% hydrophilic residues. The amino acid sequence of this protein does not display significant homologies with any of the protein sequences in the Protein Database.

The most striking features of this sequence occur between amino acid residues 200 and 350, and more particularly at positions 217 through 328. This region contains many repeats of short stretches of amino acids.

For example, the five amino acid sequence asparagine-asparagine-asparagine-isoleucine-glycine (N N N I G , using one letter code) is repeated three times consecutively at positons 227 through 241.

But perhaps the most interesting feature concerns a five amino residue sequence that displays

at least partial matches with several sequences in this region. These five residue sequence repeats begin at position 217 and continue through position 328 of Figure 2. The consensus sequence of this repeat appears to be X - glycine - asparagine - Z - glycine, or XGNZG, using one letter code. For the fifteen sequences that match this consensus sequence, X is most often phenylalanine, serine or threonine (12/15), although X can also be isoleucine, leucine and aspartic acid. Z is most often isoleucine or threonine (10/15), but is also sometimes serine, leucine or valine. Additional sequences between positions 300 and 350 display partial matches with the consensus sequence (i.e., match 2 of the 3 core residues).

The above five residue sequences are arranged, from the amino-terminus toward the carboxy-terminus, with two abutting (contiguous) XGNZG sequences that are contiguous with the three NNNIG sequences that are themselves contiguous to eight contiguous XGNZG sequences. A gap of about seventeen residues follows, that is itself followed by three contiguous XGNZG consensus sequences. Another gap of five residues ensues that abuts another two contiguous five residue XGNZG consensus sequences. Interestingly, both of those gaps contain sequences having two of the three core residues of the consensus sequence, as well as properly spaced X and Z residues.

It is further noted that this region contains a direct repeat of a fourteen amino acid residue sequence with only one mismatch (residues 295-308 and 315-328). Those sequences are shown below using one letter code:

295-308 F N S G S G N I G F G N S G
315-328 F N S G S G N I G I G N S G.

As expected, since the amino acid residue repeats of the consensus sequences are not exact, the nucleotide sequences in this region are not exact repeats. This observation suggests that recombinational processes such as an unequal crossing over may not play a role in causing rapid evolutionary changes in this region as is often observed for highly repeated nucleotide sequences.

The remainder of this protein sequence does not display any other particularly striking features.

The highly repetitious nature of the 517 residue protein is reminiscent of the repeated structures found in the major coat proteins of the sporozoite stage of the malaria parasite [Nussenzweig et al., (1985), Cell, 42:401-403]. These circumsporozoite or CS proteins are 40-60 KD proteins located on the membrane of the infectious sporozoite and contain a strongly immunodominant epitope that reacts with most of the anti-sporozoite antibodies found in polyclonal antisera as well as all of the monoclonal antibodies raised against the sporozoite stage. The central region of these proteins contains 20-40 tandemly arranged repeats of a 11-12 amino acid sequence.

In Plasmodium falciparum, the immunodominant epitope is contained within three consecutive repeats of the sequence asparagine-alanine-asparagine-proline (NANP; which is repeated 37 times in one isolate) and antibodies directed against this 12-residue repeat can provide immunologic protection against infection with the malaria parasite. The sequence of the repeat differs in the various species

of this parasite and the number of repeats can vary within different isolates of the same species. The similarity of the repeated nature of the CS protein and that of the 517 amino acid residue *M. tuberculosis* protein raises the interesting possibility that the repeated sequences in the 517 residue protein might play some role in the immune response to mycobacteria.

3. Expression of the 517 Protein

Although the 517 protein was not expressed using the before-described recombinant construct, that protein was expressed in *E. coli* using a recombinant expression vector designed specifically for its expression. That recombinant expression vector was constructed as follows, using the base pair numbering of Figure 2. It is to be understood that the DNA sequence of interest here is that shown in the lower of the two DNA sequences depicted, and that sequence, is read from right to left and in the direction from 5'-end to 3'-end, although the sequence position numbers are read from left to right and in the direction from 5'-end to 3'-end for the upper sequence.

The double stranded DNA sequence of Figure 2 was cleaved with endonuclease PvuII to provide a fragment that extends from position 3511 to position 4019 (509bp). That fragment was ligated into the SmaI site of the pUC19 vector to form intermediate I. Two orientations were possible for ligation of the PvuII fragment in the vector. Proper orientation was determined by usual methods such as isolation of several insert-containing clones and preparation of restriction maps of the DNA from those clones. For example, a BglI fragment from a clone having the PvuII DNA fragment in the proper orientation contains

about 1500 bp, whereas a BglI fragment from a clone having an improperly oriented PvuII fragment contains only about 1300 bp. Intermediate I was introduced into E. coli to propagate the vector DNA.

The propagated DNA of intermediate I was thereafter cleaved with endonucleases NotI (position 3603) and SalI (in the pUC19 polylinker site). The resulting NotI-SalI fragment was discarded, whereas the remainder of the DNA of Intermediate I was retained.

A further sample of the DNA sequence of Figure 2 was cleaved with endonucleases NotI (position 3603) and SalI (position 2202) to provide a NotI-SalI fragment that was ligated into the appropriate sites of the retained Intermediate I DNA to form a second pUC19-derived vector denominated Intermediate II. That vector contained the complete 517 protein DNA sequence, and was propagated further in E. coli.

The propagated DNA of Intermediate II was collected and cleaved with endonucleases EcoRI and HindIII at their respective sites in the 517 protein gene and in the polylinker of pUC19. The resulting EcoRI-Hind III fragment that contained the 517 protein DNA was thereafter collected and ligated into those respective sites in the polylinker of plasmid vector pKK223-3 to form Intermediate III that contained the carboxy-terminal portion of the gene. Intermediate III was cloned in E. coli JM105. (pKK223-3 and JM105 are available from Pharmacia Fine Chemicals, Piscataway, NJ.)

A further sample of the DNA of Intermediate II was cleaved with EcoRI alone to excise a portion of that DNA from a position in the polylinker to position 2969 in the 517 protein. The resulting

EcoRI fragment containing DNA that codes for the amino-terminal portion of the 517 protein was collected, and was thereafter ligated into the single EcoRI site of Intermediate III to form the expression vector that contains the entire 517 protein gene. That vector was also cultured in E. coli JM105 as a replication/expression medium.

It is noted that two orientations were possible for ligation of the EcoRI fragment in the expression vector. Proper orientation was determined by usual methods such as isolation of several insert-containing clones and preparation of restriction maps of the DNA from those clones. For example, a KpnI-HindIII fragment from a clone having the EcoRI DNA fragment in the proper orientation contains about 2000 bp, whereas a KpnI-HindIII fragment from a clone having an improperly oriented EcoRI fragment contains only about 800 bp.

Expression of a recombinant protein from vector pKK223-3 is inducible with IPTG, and the induced recombinant protein is expressed as the protein itself, and not as a fusion product. The resulting E. coli cells were thus grown and then induced with IPTG, as discussed elsewhere herein.

The expressed protein was produced in a relatively large amount and could be readily identified in an SDS-PAGE gel from a lysate of the E. coli cells. The 517 protein had an apparent M_r of about 55,000 daltons in SDS-PAGE, as expected.

The expressed 517 protein can also be collected and purified, as with an affinity column made from Sepharose 4B (Pharmacia) to which antibodies raised to one or more of the 517 protein-related peptides are bound via the cyanogen bromide activation technique, or by ammonium sulfate

precipitation, followed by DEAE-cellulose chromatography.

H. Recombinants and Vectors

The present invention thus contemplates the purified recombinant 540 protein and 517 protein, as well as those recombinant fusion proteins that also include all or a portion of another molecule such as beta-galactosidase fused to the amino-terminus of those proteins. Each of those recombinant proteins is useful for inducing the production of antibodies that immunoreact with those respective molecules as obtained from M. tuberculosis itself or from cells infected with that mycobacterium. Methods of preparing such antibodies are well known in the art and are similar to the methods utilized for the peptides of this invention as described hereinafter.

The purified recombinant 540 amino acid residue protein or its fusion proteins when present in an effective amount in an inoculum are also useful in a DCH assay, as described before. Those proteins are also useful in diagnostic methods and kits useful for assaying for the presence of infection by M. tuberculosis.

Nucleotide sequences are also contemplated, as are non-chromosomal plasmid vectors useful for propagating those DNA sequences and expressing the protein products coded for by those sequences.

A nucleotide sequence of this invention consists essentially of one of the before-described sequences. Thus, a nucleotide sequence of the invention excludes additional nucleotides that affect the basic and novel characteristics of a nucleotide sequence that codes for the 540 protein or the 517 protein.

A nucleotide sequence of the invention can include one or more transcriptional promoter sequences operationally linked to the sequence adjacent to the 5'-end thereof. Where translation of the DNA and protein expression are desired, the DNA also includes a translation initiating codon (ATG) and a translation terminating codon (TAA or TAG or TGA), each operationally linked adjacent to the 5'-end and 3'-end, respectively, of the sequence, with the translation initiating codon being located between the promoter sequence and the 5'-end.

A DNA sequence that codes for all or a portion of another molecule can also be included in the DNA molecule so that the translated (expressed) proteinaceous molecule is a fusion protein that includes an amino acid residue sequence of all or a portion of that other molecule fused (linked by a peptide bond) to the expressed 540 protein or 517 protein. An exemplary fusion polypeptide is the fusion protein molecule discussed herein that contains a portion of the beta-glactosidase molecule fused to the amino-terminus of the 540 amino acid residue protein.

All of the nucleotide sequences shown in Figure 2 can be present so long as an enumerated DNA molecule remains replicable, where only replication is desired. Where replication and translation (proteinaceous molecule expression) are desired, those nucleotide sequences are present so long as the DNA molecule remains replicable and the proteinaceous molecule containing the amino acid residue sequence of 540 protein or 517 protein expressed exhibits immunological cross-reactivity with the antibodies raised to an appropriate peptide described herein. In more preferred practice, only those base pairs needed for expression of a desired protein are utilized.

A non-chromosomal, plasmid vector for propagation and expression of a desired DNA nucleotide sequence as defined herein in a replication/expression medium, e.g., a unicellular organism or the like such as *E. coli*, *S. cerevisiae* or mammalian cells such as COS cells, is also contemplated. That vector comprises a replicon that is compatible with the replication/expression medium and contains therein the foreign DNA molecule (e.g., all or a portion of the sequence shown in Figure 2) to be replicated in a manner such that the vector can propagate the DNA molecule.

In addition, the non-chromosomal plasmid vector also includes those sequence components that are utilized for transcription and translation. To that end, a transcriptional promoter can be operationally linked to the DNA molecule present adjacent to the 5'-end thereof, as already noted. The transcriptional promoter can be endogenous to the vector or exogenous to the vector. A transcriptional promoter endogenous to the vector such as the *lac Z* promoter-operator utilized in the vectors derived from pUC19 or the *trp-lac* (*tac*) promoter of pKK223-3 is preferred. A translational terminator can also be operationally linked adjacent to the 3'-end of the DNA molecule in some instances, although the nucleotide sequence represented by the formula of Figure 2 contains such terminator sequences.

An initiation codon (ATG) adjacent to the 5'-end of the sequence that begins translation in a replication/expression medium is also required to be present in a vector used for expression. Such a codon can be present in a defined DNA molecule in frame, as is the case with the sequences shown in Figure 2, or can be a portion of the precursor plasmid vector nucleotide sequence.

The before-discussed transcription promoter, translation initiating and translation terminating codons are frequently parts of the non-chromosomal plasmid vector as compared to a DNA molecule of the invention. For use in expression of the proteinaceous molecule, the precursor plasmid frequently also includes a ribosome binding site (Shine-Delgardo sequence) adjacent to the 5'-end of the foreign DNA molecule and located upstream from the initiation codon, as is well known. The vector's promoter such as the lacZ and tac promoters utilized herein typically contain a ribosome binding site.

Thus, the nucleotide sequence of the plasmid vector used for expression, aside from those nucleotides needed for the replication and general vector function include, in frame and from 5'-end to 3'-end, a ribosome binding site operationally linked adjacent to the 5'-end of a transcription promoter; that promoter operationally linked to the 5'-end of the translation initiating codon; that codon operationally linked to the 5'-end of: (a) a sequence of a portion of another molecule that is expressed as a fusion protein with the desired protein, or (b) a foreign DNA molecule of this invention; where (b) is present, that sequence is operationally linked to the 5-end of a DNA molecule of this invention. An expression vector containing the foreign DNA molecule of this invention, (however linked adjacent to its 5'-end) also contains a translation terminating codon adjacent the 3'-end of the foreign DNA.

It is to be understood that all of the DNA sequences of the vector must be compatible with the replication/expression medium utilized for replicating the DNA, and more preferably for expressing a product coded for (encoded by) a DNA molecule of this invention.

It is also to be understood that the before-recited signal sequences of the useful vector can be supplied to that vector by the foreign DNA or by a precursor to the final vector. For example, the translation initiation and termination codons in the expression vector for the 517 protein are provided by the foreign DNA, whereas the promoter and ribosomal binding site sequences are provided by the precursor plasmid.

A vector of the invention is at least capable of replicating (propagating) a DNA molecule of the invention. More preferably, the vector is capable of not only replicating a DNA molecule, but is also capable of expressing or translating the genomic information of that DNA into a recombinant protein molecule that is immunologically similar to the 540 protein or the 517 protein; i.e., will induce cross-reactive antibodies.

A non-chromosomal plasmid vector of this invention need not be limited to those vectors useful for replication and translation (expression) in E. coli as host replication/expression medium. Substantially any vector useful for replicating (propagating) and expressing a DNA sequence can be utilized for replicating the DNA, e.g. in mammalian or eukaryotic cells.

A wide range of such vectors is commercially available as are appropriate host replication media. Exemplary vectors, both plasmids and bacteriophages and hosts are available from the American Type Culture Collection of Rockville, MD, and are listed in its CATALOGUE OF BACTERIA, PHAGES AND rDNA VECTORS, sixteenth ed., 1985. In addition, plasmids, cosmids and cloning vectors are listed as being available in catalogues from Boehringer Mannheim Biochemicals of Indianapolis, IN; Bethesda Research

Laboratories, Inc. of Gaethersberg, MD, and New England Biolabs, Inc. of Beverly, MA.

I. Peptides

Another aspect of the present invention relates to a peptide that consists essentially of an amino acid residue sequence that corresponds substantially to a portion of the 540 or the 517 protein sequence. Such a peptide contains 5 to about 40 amino acid residues, and more preferably about 10 to about 20 amino acid residues that correspond substantially in sequence to a protein of either the 540 amino acid residue protein or the 517 amino acid residue protein that are coded for by the DNA sequence shown in Figure 2.

A useful peptide most preferably contains only those amino acid residues that are identical or homologous to (conservative substitutions for) residues present in a sequence of either of the two above proteins. Additional residues of substantially any length can also be present at either or both termini of the peptide. However, any additional residues must not interfere with the activity of the peptide, as discussed hereinafter, and therefore, a peptide of this invention is said to "consist essentially" of an enumerated sequence. For example, a peptide of the invention is free of immunosuppressing sequences. In addition, if additional residues are present, and together with an above peptide correspond substantially in sequence to further portions of the same protein to which the sequence of the peptide substantially corresponds, the resulting peptide is of a molecular weight less than that of the naturally occurring 540 or 517 proteins, respectively.

A peptide of this invention is useful, inter alia, for inducing the production of antibodies in a

laboratory mammal such as a mouse or rabbit. Those induced antibodies immunoreact with the inducing peptide as well as with the protein to which the peptide sequence substantially corresponds when that protein is in an SDS-denatured form as in a Western blot analysis subsequent to SDS-PAGE analysis.

Thus, the anti-peptide antibodies can be used in solid phase assays for the detection of the presence of an antigen that is the 540 protein or the 517 protein of M. tuberculosis. In this instance, the assayed sample such as sputum provides the antigen that is affixed to the solid phase matrix to form the solid support. An aqueous composition containing the anti-peptide antibodies or their idiotypic portions (binding site-containing portions) is admixed, maintained and separated from the solid phase as previously discussed for the presence of anti-65KD protein antibodies. The presence of bound anti-peptide antibodies is thereafter assayed to determine the presence of the M. tuberculosis antigen in the sample, following the broad admixture, separation and analysis steps previously described. Whole antibodies and their idiotype-containing portions such as Fab and $F(ab')_2$ portions are collectively referred to as paratoic molecules.

In exemplary studies, antibodies (paratopic molecules) were raised in New Zealand white rabbits to both the amino-terminal and carboxy-terminal polypeptide sequences (Peptides 1 and 54 of Table 4, hereinafter) of the 540 protein. Varying dilutions of pure M. tuberculosis cultures were bound to the walls of microtiter plates to form a solid support and one or the other of the two aqueous anti-peptide antibody preparations was admixed with the solid support to form a solid/liquid phase admixture. After maintaining the solid/liquid phase admixture

for a time period sufficient for the anti-peptide antibodies to bind to the mycobacterial antigens present, the phases were separated. The solid phase was rinsed to assure removal of unbound anti-peptide antibodies. The presence of anti-peptide antibodies bound to the solid support was thereafter determined by standard methods.

As a result of those studies, it was determined that the presence of a mycobacterial antigen could be detected at a concentration of about 10^5 organisms per well. Sputum samples from persons with active infections of M. tuberculosis typically contain about 10^5-10^6 organisms in the volume of a sample utilized in the study. Thus, anti-peptide antibodies raised to a peptide of this invention such as those raised to Peptides 1 and 54 can be utilized to detect mycobacterial antigens present at a level found in a clinical environment.

Antibodies were similarly raised to the immunologically active recombinant, fusion, 540 protein produced by pTB22, and a similar antibody binding study was carried out. The results were generally similar to those discussed above, except that this assay was somewhat more sensitive, presumably as a result of the polyclonal character of the induced antibodies.

In addition to the above assays for mycobacterial antigens, several additional immunoassays can be carried out using antibodies induced (a) by the previously-mentioned 540 protein, or more particularly (b) by an immunologically active portion thereof such as the fusion protein produced by pTB22, a fusion protein that contains a peptide sequence of Tables 2 or 4 fused to a portion or all of another molecule such as beta-galactosidase, or by a peptide of Tables 2 and 4. Such additional

immunoassays are well known in the art and include, for example, double antibody, "sandwich", assays, and competition assays as where a peptide or other antigen described herein competes for the antibodies with a mycobacterial antigen in the assayed sample.

In each of the immunoassays, a sample to be assayed for the presence of a mycobacterial 65KD cell wall protein-a antigen is admixed in an aqueous medium with paratopic molecules raised to the 540 protein, or more particularly to an immunologically active portion thereof. The resulting admixture is maintained for a time period sufficient for the paratopic molecules to immunoreact with mycobacterial antigens present in the admixed sample to form an immunoreactant. The presence, and usually the amount, of immunoreactant formed is determined.

The anti-peptide paratopic molecules can themselves contain a label. Preferably, however, a second label-containing reagent is utilized that reacts with the bound paratopic molecules such as whole anti-peptide antibodies. The peroxidase-conjugated goat-anti-mouse antibodies utilized herein are exemplary of such reagents.

A solid phase assay kit that utilizes the anti-peptide antibodies or other paratopic molecules induced by an immunologically active portion of the 540 protein is also contemplated herein for clinical use of the before-described method. Here, the kit contains at least a solid phase matrix to which the assayed-for antigen of the sample or antibodies can be affixed in one package and a preparation of anti-peptide or anti-immunologically active 540 protein portion paratopic molecules that immunoreact with the 540 (65KD) protein or the 517 protein in a second package. Additional packages of reagents similar in type and function to those previously mentioned can also be included.

For inducing paratopic molecules such as whole antibodies, a useful peptide is typically linked to an antigenic carrier molecule such as keyhole limpet hemocyanin (KLH) as a conjugate, the conjugate is thereafter dispersed in a physiologically tolerable diluent as an inoculum, and the inoculum is injected into the laboratory mammal using well known procedures. The inoculated animal is maintained and given booster injections as required, until a desired antibody titer to the inducing peptide is achieved. The mammal's antibody-containing serum is thereafter obtained, purified as desired, and utilized in a diagnostic assay such as an SDS-PAGE/Western blot for the presence of a substantially corresponding protein.

The word "inoculum" in its various grammatical forms is used herein to describe a composition containing an amount of peptide conjugate, peptide polymer (as described hereinafter), 65KD protein or recombinant protein sufficient for a described purpose that is dissolved or dispersed in an aqueous, physiologically tolerable diluent. Exemplary diluents are well known and include water, physiological saline, phosphate-buffered saline, Ringer's solution, incomplete Freund's adjuvant and the like.

Inocula can contain varying amounts of a preferred peptide or polymer, depending upon its use.

Where paratopic molecules are to be formed or an inoculum is otherwise to be used as a vaccine, about 100-500 micrograms of peptide or peptide polymer are used per injection into laboratory animals such as mice, rabbits or guinea pigs. Larger amounts are utilized for larger mammals, as is known. Similar amounts of peptide or polymer are utilized for <u>in vivo</u> DCH assays.

Smaller amounts of antigenic peptide or antigenic peptide polymer are utilized for *in vitro* stimulation assays. Using 200 microliters (ul) of total volume and about 1-2x10^4 PBMC and about 1x10^5 antigen-presenting cells, concentrations of about 0.1 to about 50 micrograms of antigen per milliliter are useful.

Exemplary procedures for the chemical synthesis of a useful peptide as well as preparation of a conjugate and use of the conjugate to raise antibodies can be found in U.S. Patent No. 4,636,463, No. 4,599,231, No. 4,599,230, No. 4,545,931, No. 4,544,500, all of whose disclosures are incorporated herein by reference.

Another use for a preferred peptide of this invention is in an assay for the presence of mycobacterially-exposed (or immune); i.e., previously immunologically exposed, mononuclear cells such as T cells in a body sample containing such cells. Mycobacterially-exposed (or immune) mononuclear cells are cells that themselves have been immunologically exposed to a mycobacterial immunogen or whose progenitor cells had been so exposed to such an immunogen. Thus, a preferred peptide can be used to determine whether a mammal has been immunized against a mycobacterium or has or has had a mycobacterial infection.

In such an assay, peripheral blood mononuclear cells, and particularly T cells, from the mammal are provided. Those cells are admixed and contacted in an aqueous cell culture medium with a stimulating amount of both antigen presenting cells and a preferred peptide of the invention to form a stimulatory cell culture. The stimulatory cell culture is maintained for a period of time sufficient for immune mononuclear cells present to be stimulated

and to evidence that stimulation, usually about 18-96 hours and most usually 24-48 hours, under usual cell culture conditions. The presence of mononuclear cell stimulation is thereafter determined.

Where mononuclear cell stimulation is found, it indicates the presence in the assayed mononuclear cell population of cells which themselves were immunologically exposed to a mycobacterium or whose parental line was immunologically exposed to a mycobacterium.

As is illustrated by the results shown in Table 3, hereinbefore, the recombinant 540 protein and recombinant fusion protein containing a portion of the beta-galactosidase molecule and an immunologically active portion of the 540 protein were useful in stimulating mycobacteriologically-immune mononuclear cells *in vivo* in a DCH assay. Such molecules can be utilized in the above-described assay, and in the stimulatory assays described hereinafter in the same manner as can the peptides, and in the place of a peptide.

Mononuclear cell stimulation can be determined in a number of manners that are well known in the art, some of which are described specifically hereinafter. The cells of the mononuclear cell population that most usually are stimulated are T cells, and for that reason, the mononuclear cells will be usually referred to hereinafter as T cells. More particularly, T cells that exhibit the CD4 or T4 ($CD4^+$ or $T4^+$ antigen and those that exhibit the CD8 or T8 ($CD8^+$ or $T8^+$) antigen are the cells that are typically stimulated. Those T cells are often more generally referred to as helper and killer or cytotoxic T cells, respectively.

Exemplary manners in which T cell stimulation can be determined include (a)

proliferation as assayed by the uptake of a radiolabeled nucleoside such as [^{3}H]-deoxythymidine also referred to as [^{3}H]-Tdr, [^{3}H]-thymidine ([^{3}H]-T), (b) secretion of interferon-gamma, (c) secretion of interleukin-2 (IL-2), (d) secretion of granulocyte macrophage-colony stimulating factor (GM-CSF), (e) cytotoxicity, a phenomenon that can occur with T cells such as T4$^+$ T cells as well as with T8$^+$ cells, (f) the ability to provide an in vitro B cell helper function, (g) the ability of immune T cell clones to provide a delayed cutaneous hypersensitivity (DCH) response in vivo as described herein and in U.S. Patent No. 4,689,397 whose disclosures are incorporated by reference, and (h) the ability of immune T cell clones to provide protective immunity in vivo.

A kit is also contemplated for use with the immediately preceding assay. That kit can include a number of containers, at least one which contains a preferred peptide antigen of this invention or a polymer of such a peptide antigen whose repeating units are comprised of a "di-Cys-terminated" peptide as is described hereinafter. A mixture of two or more such preferred peptides or their polymers can also be present. A sufficient amount of a preferred peptide or peptide polymer is contained in the container to perform at least one assay using that method.

The assay kit can further include a premeasured amount of buffer or other salt for the preparation of an inoculum of the peptide or polymer upon the addition of water or other suitable aqueous medium. The inoculum can also be provided in premixed aqueous form either at the concentration for use or as an aqueous concentrate to be diluted.